Checkpoint #1

Project 10: Casey Grage & Renee Zha

In terms of what we learned about CPDB, we found police have a *lot* of allegations, a lot of allegations with victims, and often, have multiple allegations! The numbers are outlined below, of course, but the main takeaway was how shockingly high the numbers are. Beyond the data itself, this was the first time either of us used SQL, so we mostly just became acquainted with the system and learned what kind of functions the language has built in. We learned basic functions like `SELECT, COUNT, FROM, GROUP BY, ORDER BY, HAVING, WHERE,` and `NOT IN`. We found questions 1-3 to be fairly straightforward (especially after the initial crash course from Jennie). The most challenging question was no. 4. We learned quickly that it is a good check to write different queries that should yield the same output. Using this strategy, we were able to try a couple different queries until we were confident in the answer we were receiving.

1. Which identity groups filed the most complaints? For officer AND civilian complainants

   QUERY:

   SELECT gender, race, COUNT(*) FROM data_complainant GROUP BY gender, race ORDER BY COUNT(*) DESC;

```
 gender |              race               | count
--------+---------------------------------+-------
 M      | Black                           | 26004
 F      | Black                           | 24232
 M      | White                           | 11746
 F      | White                           |  6183
 M      | Hispanic                        |  4860
 M      |                                 |  4042
 F      | Hispanic                        |  3619
        |                                 |  2354
 F      |                                 |  1852
 M      | Asian/Pacific Islander          |   499
 F      | Asian/Pacific Islander          |   210
        | Black                           |    86
 M      | Native American/Alaskan Native  |    62
 F      | Native American/Alaskan Native  |    47
        | White                           |    32
        | Hispanic                        |    15
        | Asian/Pacific Islander          |     3
```

2. What percentage of unique officers have multiple allegations of any kind against them? (from “data allegations” in cpdb)

   Finding no. unique officers w multiple allegations QUERY:

   SELECT COUNT(DISTINCT officer_id) FROM data_officerallegation WHERE officer_id NOT IN (SELECT officer_id FROM data_officerallegation GROUP BY officer_id HAVING COUNT(*)=1);

```
 count
-------
 19984
(1 row)
```

   Finding no. officers w allegations QUERY: :

   SELECT COUNT(distinct officer_id) FROM data_officerallegation;

```
 count
-------
 22813
(1 row)
```

   Finding total no. of officers in database QUERY:

   SELECT COUNT(*) FROM data_officer;

```
 count
-------
 33671
(1 row)
```

Organizing the data in excel:

total no. officers	33671
no. officers w allegations	22813
no. unique officers w multiple allegations	19984

   no. unique officers w multiple allegations / no. officers w allegations = 0.875991 * 100 = **87.6%**

   no. officers w allegations / no. total officers = 0.677526655 * 100 = **67.6%**

   no. unique officers w multiple allegations / no. total officers = 0.593507 * 100 = **59.4%**

3. Which identity groups are most often victims? For officer AND civilian complainants

   QUERY:

SELECT gender, race, COUNT(*) FROM data_victim GROUP BY gender, race ORDER BY COUNT(*) DESC;

```
 gender |              race               | count
--------+---------------------------------+-------
 M      | Black                           | 36856
 F      | Black                           | 22342
 M      | White                           |  7384
 M      | Hispanic                        |  7017
 F      | White                           |  4990
 F      | Hispanic                        |  3585
 M      |                                 |  2519
        |                                 |  2286
 F      |                                 |  1485
 M      | Asian/Pacific Islander          |   511
 F      | Asian/Pacific Islander          |   229
        | Black                           |   100
 M      | Native American/Alaskan Native |    77
 F      | Native American/Alaskan Native |    43
        | White                           |    32
        | Hispanic                        |    24
        | Asian/Pacific Islander          |     7
```

4. What percentage of unique officers have multiple victim allegations against them? (from “data allegations” in cpdb)

Finding no. officers w allegations w a victim QUERY:

SELECT COUNT(distinct officer_id) FROM data_officerallegation oa, data_victim v where v.allegation_id = oa.allegation_id;

```
 count
-------
 11814
(1 row)
```

Finding no. officers w multiple allegations w a victim QUERY:

SELECT COUNT (distinct officer_id) FROM data_officerallegation oa, data_victim v WHERE v.allegation_id = oa.allegation_id AND officer_id NOT IN (SELECT officer_id FROM

data_officerallegation oa, data_victim v WHERE v.allegation_id = oa.allegation_id GROUP BY officer_id HAVING COUNT(*)=1);

```
 count
-------
 10625
(1 row)
```

Finding total no. of officers in database QUERY:

SELECT COUNT(*) FROM data_officer;

```
 count
-------
 33671
(1 row)
```

Organizing the numbers in excel:

total no. officers	33671
no. officers w allegations w a victim	11814
no. unique officers w multiple allegations w a victim	10625

no. unique officers w multiple victim allegations / no. officers w victim allegations = 0.89935 *100 = **89.9%**

no. officers w victim allegations / no. total officers = 0.350865 * 100 = **35.1%**

no. unique officers w multiple victim allegations / no. total officers = 0.31555 * 100 = **31.6%**